Plant Molecular Biology **34:** 809–814, 1997.

Short communication

Identification of the tobacco and *Arabidopsis* homologues of the pollen-expressed LAT59 gene of tomato

Rima Kulikauskas and Sheila McCormick*
*Plant Gene Expression Center, USDA/ARS-UC Berkeley, 800 Buchanan St., Albany, CA 94710, USA (*author for correspondence)*

Received 8 January 1997; accepted in revised form 16 April 1997

Key words: allergen, *Arabidopsis thaliana*, gene expression, intron, pectate lyase, pollen-specific

Abstract

We describe the complete genomic sequences for the tobacco and *Arabidopsis* homologues of tomato LAT59, a previously described member of a family of pectate lyase-like genes. Translation of the tobacco gene, Nt59, predicts a protein with 93.5% overall amino acid similarity to LAT59. Nt59 has two introns whose positions are exactly conserved with the two introns of LAT59. Both LAT59 and Nt59 are specifically expressed in pollen and their promoter and 5′-UTR sequences are highly similar. Furthermore, two promoter elements shown to be important for pollen expression of LAT59 are conserved in the Nt59 promoter. The *Arabidopsis* homologue, At59, was found by examination of four candidates. At59 has 72.6% amino acid similarity to LAT59 and the position of one of its two introns is conserved with one of the LAT59 introns. At59 is also pollen-expressed and although its promoter sequence is quite different from the Nt59 and LAT59 promoters, the two promoter elements are somewhat conserved.

LAT56, LAT59 and 9612 are members of a growing family of pectate lyase-like genes that have been described in numerous plant species [3, 29]. LAT56 and LAT59 are distinct genes preferentially expressed in tomato anthers and pollen; the 9612 gene shows a high degree of similarity to these sequences, yet it is pistil-expressed. This family also includes the ragweed pollen allergens AmbaI.1, AmbaI.2, and AmbaI.3 [17, 18], G10, a probable tobacco homologue of LAT56 [19], Zm58.2 from maize and LMP131 from lily, which are most similar to LAT59 [25, 12], and the Japanese cedar pollen allergen *Cry j* I [22]. These genes have significant amino acid similarity to pectate lyases of the plant pathogen *Erwinia* [27, 29]. The amino acid similarities are especially high in regions I and II, which are conserved among all pectate lyases described in *Erwinia* [23].

Although all these plant sequences are obviously similar to pectate lyases, only one of the plant proteins, *Cry j* I, has thus far been shown to have pectate lyase activity [24]. Ragweed allergens are surface-localized proteins [13] and we think that LAT56 and LAT59 proteins are also on the surface because they are secreted and glycosylated [4]. We have suggested [4, 29] that the plant members of this family might play roles in cell wall deposition as the pollen tube grows through the pistil. Sequence and structural alignments [10] of the existing pectate lyase family members have been useful in delimiting invariant residues likely to be important for protein structure or function; additional plant sequences may help to refine analyses of the these proteins.

In addition to determining the function of the LAT proteins, our group has studied the regulation of pollen-expressed genes. We previously showed that the LAT52 and LAT59 promoters direct pollen-specific expression in tobacco and *Arabidopsis* [26] and we identified *cis* elements in the promoters of the LAT52,

The nucleotide sequence data reported will appear in GenBank, EMBL and DDBJ Nucleotide Sequence Databases under accession numbers U85646 and U95924 (Nt59), U83619 (At59), U83620 (A1), U83621 (A10) and U83622 (A11).

LAT56, and LAT59 genes that are important for pollen-specific expression [7, 27]. In order to determine if these promoter elements were conserved in homologues, we used the LAT59 gene and tobacco and *Arabidopsis*. The LAT59 promoter was functional in these species [26] and additionally, LAT59-like members of the gene family have a distinctive N-terminal extension [12, 25, 29] that would potentially make it easier to identify homologous sequences. Tomato and tobacco are in the same plant family (Solanaceae) in the order Polemoniales and are estimated to be separated in evolution by only about 15 million years [11]. *Arabidopsis thaliana* is a member of the Papaverales; the precise evolutionary relationship between the Polemoniales and the Papaverales is not known [2], but the separation is presumably less than the 200 million years estimated to represent the divergence of monocots and dicots [30]. Here we report the isolation of genomic clones from tobacco and *Arabidopsis* which encode homologues of the LAT59 protein.

We used homologous primers corresponding to the N-terminal half of the LAT59 protein (amino acids 1–123) for PCR on tobacco genomic DNA, and used the resulting cloned fragment as the probe to screen 200 000 pfu of a purchased *Nicotiana tabacum* genomic library (Clontech) using standard methods [21]. A 3 kb and a 0.5 kb *Bam*HI fragment from the same positive lambda clone were subcloned and sequenced. Although the lambda clone lacked the 3′ end of the gene, we were able to obtain this sequence using the Perkin Elmer Gene Amp Thermostable rTth Reverse Transcriptase RNA PCR kit with tobacco pollen total RNA. A poly-T primer with a BamHI cleavage site at the 5′ end served as the 3′ primer in the reverse transcriptase reaction, while PCR was performed with the addition of a gene specific 5′ primer annealing in the third exon. RT-PCR was also used to obtain a product excluding the sequence of the first intron using primers which annealed in the first and second exons. The position of the second intron was inferred from the reading frame and stop codon sites.

The complete Nt59 gene has 86.2% overall nucleotide identity with LAT59 in the coding region [9]. The gene is interrupted by two introns of 105 and 72 bp whose positions are precisely conserved with those of LAT59 (Figure 1A). Translation of Nt59 predicts a peptide 446 amino acids long, which has 93.5% similarity and 87.9% identity with LAT59. Furthermore, Nt59 has the N-terminal extension present in LAT59, but not in LAT56, another pollen-specific pectate lyase-like gene [4, 29].

The northern blot autoradiogram in Figure 2A shows that, like LAT59, Nt59 is strongly expressed in mature pollen. Expression is also detected in stamens, presumably due to gene expression in the pollen within the anthers. We have previously shown that LAT59 is not expressed in anthers washed free of pollen [26]. Nt59 expression was not detected in leaves (data not shown), sepals, petals, stigma and styles, or ovaries. The Nt59 mRNA transcript is about 1.8 kb.

To look for conservation of cis-acting elements between tomato and tobacco for this gene, we compared their promoters. LAT59 and Nt59 have 78.4% identity for ca. 420 nucleotides of the promoter. This region includes the complete leader sequences for both genes. Extension of primers [1] complementing bp +4 to +39, or −99 to −64, each identified two transcription initiation sites for Nt59 (data not shown). The two located start sites were mapped to −207 (G) and −206 (A) nucleotides relative to the initiator methionine, resulting in untranslated leaders similar in length to that of LAT59 (236 nt, [27]). There is putative TATA box, 'TATTA', beginning 38–39 nucleotides upstream of the extension products of Nt59. The 56/59 box, identified in the LAT56 and LAT59 promoters and shown to be important for high levels of gene expression [27], is conserved in Nt59 except for one base pair (Figure 1B). In addition, the 20 bp sequence identified as essential for pollen-specific expression of LAT59 [7] is well conserved (16/20 nt) in the tobacco gene.

To find the homologue of LAT59 in *Arabidopsis*, we began again by making a suitable probe to screen a genomic library. Using degenerate primers to pectate lyase conserved regions, we amplified products from *Arabidopsis* genomic DNA. The primer sequences corresponding to the conserved amino acids CWRCDPNW and NDYTHWNM were: 59CR1, with a 5′ *Bgl*II site (5′-GAAGATCTG-(T/C)TGGAG(A/G)TG(T/C)GA(T/C)CCIAA(T/C)T-GG-3′); and 59CR2, with a 5′ *Eco*RI site (5′-GGA-ATTCAT(A/G)TTCCA(G/A) TGIGT(A/G)TA(A/G)-TC(A/G)TT-3′). This PCR produced DNA products of ca. 800, 900, and 975 bp. After gel purification, cloning, and sequencing, we identified four unique candidates: A1 (566 bp), A2 (823 bp), A10 (795 bp) and A11 (896 bp) (Figure 1A). Note that none of the four has an intron conserved with the second intron of LAT59, nor do their intron positions correspond to the intron positions in LAT56 [29]. Clone A1 is shorter than the other three because of an internal *Bam*HI site, which resulted in the loss of the 3′-prime section during cloning. When used to search GenBank,

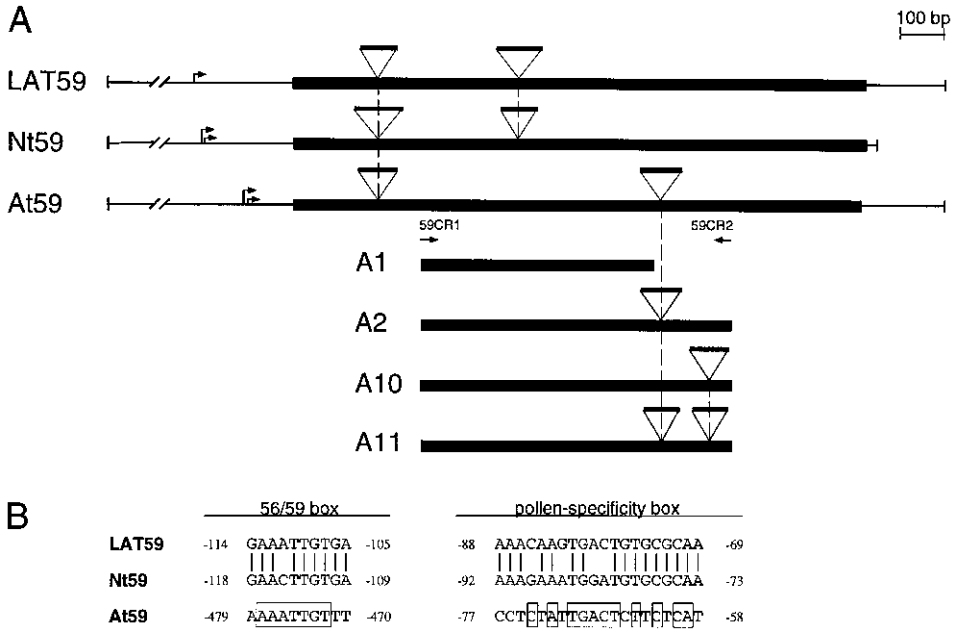

Figure 1. Comparison of the gene structures of LAT59, Nt59, and At59. A. Graphical representation of LAT59, Nt59, and At59 genomic sequences and the four *Arabidopsis* PCR products. Coding regions are represented by solid boxes and intron positions are indicated by triangles. Dashed lines indicate conserved intron positions. Bent arrows indicate transcription initiation sites. Arrows labeled 59CR1 and 59CR2 mark the positions of the degenerate primers. B. Alignments of promoter elements important for pollen-specific expression. Nucleotide positions listed are relative to the beginning of the longest leader of each gene. Boxed nucleotides in At59 match nucleotides in *either* LAT59 or Nt59.

all showed highest sequence similarity to two tomato pectate-lyase family members: the pistil-expressed 9612 gene, and LAT59. Only A1 corresponds to an EST clone (Q-ATTS4502) in the TIGR Arabidopsis Database (www.tigr.org/tdb/at/at.html). The A2 amino acid sequence is the most similar to LAT59 and Nt59, ca. 81% in each case. A1 is the least similar to Nt59 (ca. 77%), while A1 and A10 are the least similar to LAT59 (ca. 78% each). Among the *Arabidopsis* clones, A1 and A11 are the most similar (ca. 94%); A2 and A10 are the least similar (ca. 73%).

The results of Southern analysis for the four *Arabidopsis* clones are shown in Figure 2B. Each clone hybridizes intensely to a single, yet different, fragment demonstrating that each corresponds to a single copy gene. The second, minor hybridizing band on the A1 blot most likely corresponds to A11, which is 94% similar at the amino acid level and has a similar banding pattern.

Tissue specific expression patterns of the four *Arabidopsis* LAT59-like genes were determined by Northern analysis. Figure 2C shows that only A2 is expressed in pollen, while A1, A10, and A11 are expressed in flowers. None of the four are expressed in leaves. The four genes produce mRNAs of distinct sizes: A1 is 1.9 kb, A2 is 1.85 kb, A10 is 1.8 kb, and A11 is 2.3 kb.

Based on its higher degree of sequence similarity to LAT59 and especially its high expression in pollen, clone A2 was chosen to screen 200 000 pfu of an *A. thaliana* (Landsberg *erecta*) genomic library (Clontech) [21]. The 3 kb complete genomic clone obtained from the screen was designated At59 (Figure 1A). The exact positions of the 85 and 97 bp introns were determined by RT-PCR: primers used to obtain this RT-PCR clone annealed in exons 1 and 3. The position of the first intron is conserved with the first introns of both LAT59 and Nt59, while the position of the second intron is unique. The At59 sequence predicts a 460 amino acid protein with ca. 73% amino acid similarity to LAT59 and to Nt59. At59 also has the N-terminal extension diagnostic for LAT59-like family members [4, 12, 25, 29].

Two transcription start sites were identified for At59 by primer extension [1] using a primer complementing bp +4 to + 36 of At59 (data not shown). Both transcripts begin with T and are 111 and 103 nucleotides upstream of the initiator codon. The sequence 'TATTA', a putative TATA box, occurs 29 nucleotides

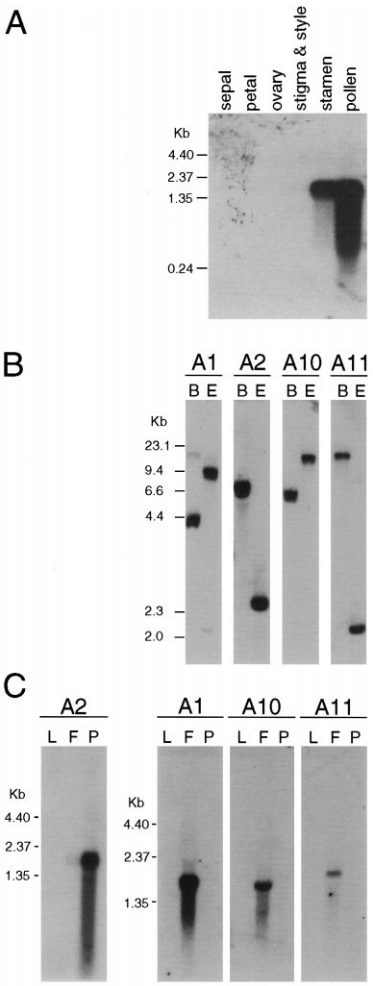

Figure 2. RNA blot and Southern Analysis of Nt59 and the four *Arabidopsis* homologue candidates. RNA, DNA, and blots were prepared as described [14, 8, 5, 21]. The tobacco northern blot was hybridized to the ^{32}P-labeled RT-PCR product which excludes the first intron, while the *Arabidopsis* northern and Southern blots were probed with complete inserts of clones A1, A2, A10, or A11. A. Tissue-specific expression of Nt59 by northern blot analysis. 10 μg of total RNA from *Nicotiana tabacum* (Samsun NN) tissues were loaded as indicated. Tobacco pollen was collected as described [16]. B. Southern blot analysis of LAT59 homolog candidates from *Arabidopsis*. Lanes B and E contained 2 μg of *Arabidopsis thaliana* (ecotype NO-O) genomic DNA digested with *Bam*HI and *Eco*RI respectively. C. Northern blot analysis of LAT59 homologue candidates from *Arabidopsis*. Each blot contained 5 μg of total RNA from *Arabidopsis* rosette leaf (L), flowers (F), and mature pollen (P). The autoradiogram shown for A2 is an overnight exposure. For A1, A10, and A11, the film was exposed for 3 days. Pollen was harvested from *Arabidopsis* plants by shaking a 50 ml conical tube full of flowers with 25 ml TE for 15 min. Two such rinses were poured through one layer of Miracloth, the pollen was pelleted in a Beckman TJ-6 centrifuge at 3000 rpm for 10 min, and resuspended in 1 ml RNA extraction buffer (4 M guanidine thiocyanate, 25 mM sodium citrate pH 7.0, 1.5% sarkosyl, 0.1 M 2-mercaptoethanol). The pollen was re-pelleted by centrifugation in an Eppendorf microcentrifuge for 2 min at 14 000 rpm. The supernatant was removed and the pollen pellet (50–100 μl) was stored at -80 °C until used.

upstream of the longer transcript. In contrast to the high similarity found between the Nt59 and LAT59 promoters, no significant sequence similarity was found when the entire 1949 bp of the cloned At59 promoter was compared to the Nt59 or LAT59 promoters. Despite this overall lack of similarity, two putative pollen boxes were located in the At59 promoter (Figure 1B). The 56/59 box in At59 shares 7 out of 10 residues with LAT59, but is more than 350 bp farther upstream than the corresponding box in the LAT59 and Nt59 promoters. The pollen specificity box shares only 11 of 20 residues, with either LAT59 or Nt59 or both but is similarly located relative to the transcription start sites.

Our primary goal in isolating the homologues of LAT59 was to determine the conservation of pollen regulatory elements previously identified for LAT59 [7, 27]; however, we became interested in the challenge of identifying homologues within gene families. Using a sequence comparison approach, we successfully isolated pectate lyase-like gene sequences from tobacco and *Arabidopsis* and were able to identify them as LAT59-like, distinguishing them from LAT56. We are confident that Nt59 is the true homologue of LAT59: supporting evidence includes its high amino acid similarity, identical exon/intron organization, pollen-specific expression, and high nucleotide identity in the 5′-UTR and in the promoter, which contains two pollen regulatory elements previously described for LAT59. Despite the greater evolutionary distance and hence greater sequence divergence between tomato and *Arabidopsis*, we also believe that At59 is the *Arabidopsis* homologue of LAT59. Evidence that At59 is a homologue includes its amino acids similarity, conservation of one intron position, pollen-specific expression, and the conservation of two regulatory elements in the promoter.

Conserved intron positions are used to support arguments for evolutionary conservation, but variable introns do not necessarily indicate unrelatedness [20]. The observation that the second intron position is unique in At59, though conserved between LAT59 and Nt59, presumably indicates a change that occured before the divergence among the Solanaceae. A similar difference in intron positions was noted when the wheat and *Arabidopsis* SBPase orthologues were compared; these genes have 98.5% amino acid sequence similarity and six conserved introns, but each also has one unconserved intron [28]. The variable intron is in the putative transit region of the gene where the protein sequence is less conserved among family members.

The promoter elements are conserved among LAT59, Nt59, and At59, and considering that the At59 promoter is quite divergent overall, the identification of these putative pollen boxes in notable. The 56/59 box located in Nt59 differs from the LAT59 version by only one base pair (a C rather than an A in the fourth position) (Figure 1B), the same single position that differs in the LAT56 gene (which has a T rather than an A). This is rather strong similarity compared to the conservation of this box between tomato LAT56 and G10, its homologue in tobacco [15, 19, and D. M. Lonsdale, personal communication], where only six residues are conserved: GAATTTGTGA for LAT56 and TAACGAGTGA for G10 (also called P1-T19). Interestingly, 7 out of 10 residues of the 56/59 box are conserved between LAT59 and At59. Since the 56/59 box acts as an expression enhancer when placed in different positions and orientations in expressions constructs [27], it is probably not significant that it is located farther upstream in the At59 promoter than in the LAT59 and Nt59 promoters. The 20 bp element shown in LAT59 to be essential for pollen specific expression [7] is 80% (16/20 bp) conserved in Nt59 but only 55% (11/20 bp) conserved in *Arabidopsis*. Conservation of the position (relative to the transcription start site) of the pollen specificity box in At59 strengthens its identification as an important promoter element, although proof would require functional analysis of the At59 promoter. Examples of orthologues with divergent promoters include the *Xenopus* and chicken *nov* genes, which share conserved genomic structures and developmental regulation but have different promoter sequences except for several conserved oligonucleotide motifs [31]. Similarly, although the chicken and rat homologs of the βB1-crystallin genes have the same intron/exon structure, lens-specific expression, protein structural motifs, and conservation of promoter elements, overall their promoters are different [6].

During our attempt to clone the LAT59 homologue from *Arabidopsis* we were faced with choosing among four candidates, and found that it was necessary to examine as many features of these candidates as possible. Although the A2 PCR fragment had the highest amino acid similarity to the corresponding region in LAT59, the other genes (A1, A10, and A11) were also very similar. It was necessary to test A2 for pollen-specific expression, and to isolate the complete At59 gene to determine whether it contained an N-terminal extension (as in LAT59), since this feature is not present in all pectate lyase family members (such as LAT56). These strategies could similarly be applied to determine whether A1, A10, or A11 might be homologous to 9612, the tomato pistil-expressed pectate lyase gene [3].

Acknowledgements

We thank Yoram Eyal for designing degenerate primers and obtaining clones A1 and A2, Jaap Kooijman for sequencing clones A1 and A2, Keith Hamby for screening the tobacco genomic library, Daphne Preuss and Shurong Huang for advice on Arabidopsis pollen isolation, and David Hantz for excellent greenhouse assistance. We thank the members of our lab for helpful discussion. This work is supported by USDA CRIS grant 5335-21000-008-OOD.

References

1. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K: Current Protocols in Molecular Biology, pp. 4.8.1–4.8.5. Wiley, New York (1994).
2. Benson L: Plant Classification. D.C. Heath, Lexington (1979).
3. Budelier KA, Smith AG, Gasser CS: Regulation of a stylar transmitting tissue-specific gene in wild-type and transgenic tomato and tobacco. Mol Gen Genet 224: 183–192 (1990).
4. Dircks LK, Vancanneyt G, McCormick S: Biochemical characterization and baculovirus expression of the pectate lyase-like LAT56 and LAT59 pollen protiens of tomato. Plant Physiol Biochem 34: 509–520 (1996).
5. Doyle JJ, Doyle JL: Isolation of plant DNA from fresh tissue. Focus 12: 13–15 (1990).
6. Duncan MK, Roth HJ, Thompson M, Kantorow M, Piatigorsky: Chicken βB1 crystallin: gene sequence and evidence for functional conservation of promoter activity between chicken and mouse. Biochim Biophys Acta 1261: 68–76 (1995).
7. Eyal Y, Curie C, McCormick S: Pollen specificity elements reside in 30 bp of the proximal promoters of two pollen-expressed genes. Plant Cell 7: 373–384 (1995).
8. Fourney RM, Miyakoshi J, Day RS (III), Paterson MC: Northern blotting: efficient RNA staining and transfer. Focus 10: 5–7 (1988).
9. Genetics Computer Group, Program manual for the GCG Package, version 7, April 1991. GCG, Madison, WI (1991).
10. Jurnak F: Functional implications of structure-based sequence alignment of proteins in the extracellular pectate lyase super family. Plant Physiol 107: 963–976 (1995).
11. Kawagoe Y, Kikuta Y: Chloroplast DNA evolution in potato (*Solanum tuberosum* L.). Theor Appl Genet 18: 13–20 (1991).
12. Kim S-R, Finkel D, Chung Y-Y, An G: Abundance patterns of lily pollen cDNAs: characterization of three pollen-preferential cDNA clones. Sex Plant Reprod 7: 76–86 (1994).
13. Knox RB, Heslop-Harrison J, Reed C: Localization of antigens associated with the pollen wall by immunofluorescence. Nature 225: 1066–1068 (1970).

14. Logemann J, Schell J, Willmitzer L: Improved method for the isolation of RNA from plant tissues. Anal Biochem 163: 16–20 (1987).
15. Lonsdale DM, Allen RL, Belostotsky D, Ghose TK, Harvey AJ, Rogers HJ, Tebbut SJ, Trick M: An analysis of the relative activities of a number of promoter constructs from genes which are expressed during late pollen development as determined by particle bombardment. Plant Cell Rep 15: 154–158 (1995).
16. McCormick S, Yamaguchi J, Twell: Deletion analysis of pollen-expressed promoters. In vitro Cell Devel Biol 27P: 15–20 (1991).
17. McCormick S: Molecular Analysis of male gametogenesis in plants. Trends Genet 7: 298–303 (1991).
18. Rafner T, Griffith IJ, Kuo M-C, Bonds JF, Rogers BL, Klapper DG: Cloning of *Amb a* I (Antigen E), the major allergen family of short ragweed pollen. J Biol Chem 226: 1229–1236 (1991).
19. Rogers HJ, Harvey A, Lonsdale DM: Isolation and characterization of a tobacco gene with homology to pectate lyase which is specifically expressed during microsporogenesis. Plant Mol Biol 20: 493–502 (1992).
20. Rogers JH: The role of introns in evolution. FEBS Lett 268: 339–343 (1990).
21. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
22. Sone T, Komiyama N, Shimizu K, Kusakabe T, Morikubo K, Kino K: Cloning and sequencing of cDNA coding for *Cry j* I, a major allergen of Japanese cedar pollen. Biochem Biophys Res Commun 199: 619–625 (1994).
23. Tamaki SJ, Gold S, Robeson M, Manulis S, Keen NT: Structure and organization of the pel genes from *Erwinia chrysanthemi* EC16. J Bact 170: 3468–3478 (1988).
24. Taniguchi Y, Ono A, Sawatani M, Nanba M, Kohno K, Usui M, Kurimoto M, Matuhasi T: *Cry j* I, a major allergen of Japanese cedar pollen, has pectate lyase enzyme activity. Allergy 50: 90–93 (1995).
25. Turcich MP, Hamilton DA, Mascarenas JP: Isolation and characterization of pollen-specific maize genes with sequence homology to ragweed allergens and pectate lyases. Plant Mol Biol 23: 1061–1065 (1993).
26. Twell D, Yamaguchi J, McCormick S: Pollen-specific gene expression in transgenic plants: coordinate regulation of two different tomato gene promoters during microsporogenesis. Development 109: 705–713 (1990).
27. Twell D, Yamaguchi J, Wing RA, Ushiba J, McCormick S: Promoter analysis of genes that are coordinately expressed during pollen development reveals pollen-specific enhancer sequences and shared regulatory elements. Genes Devel 5: 496–507 (1991).
28. Willingham NM, Lloyd JC, Raines CA: Molecular cloning of the *Arabidopsis thaliana* sedoheptulose-1,7-biphosphatase gene and expression studies in wheat and *Arabidopsis thaliana*. Plant Mol Biol 26: 1191–1200 (1994).
29. Wing RA, Yamaguchi J, Larabell SK, Ursin VM, McCormick S: Molecular and genetic characterization of two pollen-expressed genes that have sequence similarity to pectate lyases of the plant pathogen *Erwinia*. Plant Mol Biol 14: 17–28 (1989).
30. Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H: Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86: 6201–6205 (1989).
31. Ying Z, King ML: Isolation and characterization of *xnov*, a *Xenopus laevis* ortholog of the chicken *nov* gene. Gene 171: 243–248 (1996).